



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Chinese Gini Coefficient from 2005 to 2012, Based on 20 Grouped Income Data Sets of Urban and Rural Residents

Citation for published version:

Chen, J, Fang, F, Hou, W, Li, F, Pu, M & Song, M 2015, 'Chinese Gini Coefficient from 2005 to 2012, Based on 20 Grouped Income Data Sets of Urban and Rural Residents', *Journal of Applied Mathematics*, vol. 2015, 939020. <https://doi.org/10.1155/2015/939020>

Digital Object Identifier (DOI):

[10.1155/2015/939020](https://doi.org/10.1155/2015/939020)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Applied Mathematics

Publisher Rights Statement:

© Chen, J., Fang, F., Hou, W., Li, F., Pu, M., & Song, M. (2015). Chinese Gini Coefficient from 2005 to 2012, Based on 20 Grouped Income Data Sets of Urban and Rural Residents. *Journal of Applied Mathematics*, 2015, [939020]. 10.1155/2015/939020

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Research Article

Chinese Gini Coefficient from 2005 to 2012, Based on 20 Grouped Income Data Sets of Urban and Rural Residents

Jiandong Chen,¹ Fuqian Fang,² Wenxuan Hou,³ Fengying Li,⁴ Ming Pu,⁵ and Malin Song⁶

¹School of Public Finance and Taxation, Collaborative Innovation Center for Financial Security, Southwestern University of Finance and Economics, Chengdu 611130, China

²School of Economics, Renmin University of China, Beijing 100086, China

³University of Edinburgh Business School, University of Edinburgh, Edinburgh EH8 9JS, UK

⁴School of Economics and Mathematics, Southwestern University of Finance and Economics, Chengdu 611130, China

⁵School of Insurance, Southwestern University of Finance and Economics, Chengdu 611130, China

⁶Anhui University of Finance and Economics, Bengbu 233030, China

Correspondence should be addressed to Fengying Li; lify0308@163.com

Received 9 November 2014; Revised 26 January 2015; Accepted 26 January 2015

Academic Editor: Jinyun Yuan

Copyright © 2015 Jiandong Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data insufficiency has become the primary factor affecting research on income disparity in China. To resolve this issue, this paper explores Chinese income distribution and income inequality using distribution functions. First, it examines 20 sets of grouped data on family income between 2005 and 2012 by the *China Yearbook of Household Surveys, 2013*, and compares the fitting effects of eight distribution functions. The results show that the generalized beta distribution of the second kind has a high fitting to the income distribution of urban and rural residents in China. Next, these results are used to calculate the Chinese Gini ratio, which is then compared with the findings of relevant studies. Finally, this paper discusses the influence of urbanization on income inequality in China and suggests that accelerating urbanization can play an important role in narrowing the income gap of Chinese residents.

1. Introduction

Several conflicts exist over the calculation of China's Gini coefficient. A literature review reveals over 30 different estimations of the Chinese Gini coefficient, all lacking a consensus. The estimations of the Gini coefficient of 1995 best exemplify this situation. While Chen [1] estimated a Gini coefficient of 0.365, in 2002, Chen and Zhou [2] used two different methods and obtained a result of 0.38392 and 0.41914. The latter results were similar to Chen's [3] and Ravallion and Chen's [4] 0.4169 and 0.415; however, these similar results were obtained using entirely different methods. Similarly, Xiang [5], Chotikapanich et al. [6], and Xu and Zhang [7] used different methods and derived similar results: 0.3515, 0.3506, and 0.3591, respectively. Hu et al. [8] and Hu [9] adopted the method of distribution function and obtained the values of 0.3761 and 0.3691. Zhao et al. [10] estimated a value of 0.445. The highest value, 0.452, was provided by Khan and

Riskin [11], which is 28.9% higher than the 0.3506 by Chotikapanich et al. [6]. Since the National Bureau of Statistics' survey data on original households has not been made public, major disagreements on the calculation of a Gini coefficient for resident income in China have persisted since the reform and opening up. The Gini coefficient is used as basic statistical data to analyze income inequality. Thus, the discrepancy in estimation methods, combined with data insufficiency, has largely limited research on China's income inequality [12, 13]. Following is a brief introduction to the existing data sources and this study's primary data source.

According to the *China Statistical Yearbook* (2012), in 2011, the household survey teams of the National Bureau of Statistics (NBS) conducted a survey on 66,000 urban households and 75,000 rural households in 7,100 villages. Thus, in comparison with others sources, NBS is a far superior source of data on resident income and regions. However, the yearbook has also been criticized. For instance, Khan and

Riskin [11] stated that the statistical data in the yearbook was too aggregated to conduct a careful and deep analysis on income inequality. Similarly, Fang et al. [14] believed that the income disparity within each group had been ignored owing to the aggregated data and the results were not accurate enough. In addition, disputes exist over the yearbook's income standard. According to Li and Luo [13], the hidden subsidies of urban residents were much higher than those of rural residents. In other words, the actual income gap between urban and rural residents was larger than that depicted by the yearbook. Sutherland and Yao [15] pointed out that NBS did not fully consider factors such as welfare disparity between urban and rural residents, cost of living in different regions, rapid expansion of urban areas, and the large number of migrant workers. To this effect, various scholars hold different opinions on the income gap between urban and rural areas. While Sicular et al. [16] estimated a small income gap, Li and Luo [17] found a dramatic and widening gap.

The Chinese Household Income Projects (CHIPs) measure income distribution using data from surveys conducted on households in selected provinces and cities for 1988, 1995, 2002, and 2007. The data were gathered by the research group at the Institute of Economic Research of Chinese Academy of Social Science. Using the adjusted data from the 1988 and 1995 surveys, Khan and Riskin [11] and Zhao et al. [10] analyzed the income distribution of residents in China. Wang [18] used the data from the 1988 and 1995 surveys to examine income mobility. *Research on Income Distribution in China*, edited by Li et al. [19], was based on the 2002 survey, while Shi Li and Sui Yang referenced the 2007 data, whose conclusions attracted much attention. However, the sample size of CHIPs is much smaller than that of NBS; for example, CHIPs 1995 samples included 14,929 households from 19 provinces, while NBS was 10,286 households for the same year.

The China Health and Nutrition Survey (CHNS) is a collaborative project by the Population Research Centre at the University of North Carolina, the National Institute of Nutrition and Food Safety, and the China Disease Prevention and Control Centre. In addition to income, the survey uses data on residents, nutrition, health, adults, children, and communities, among others, to analyze income distribution. Shi et al. [20] examined CHNS' 1997 data on rural and urban income and indicated that NBS provided marginally higher income distribution than that in CHNS. Wei [21] analyzed 1993 data for rural areas to explore factors influencing nonagricultural employment and salary. Wang [22] used 1989–1997 data to study income mobility [22] and 1989–2006 data to examine fairness in income opportunities (2012). Moreover, using the 1989–1997 data, Zhang et al. [23] analyzed changes in income distribution, and Zhu and Luo [24] studied the relationships between income inequality, poverty, and economic growth in China based on the same data sources. Between 1989 and 2009, CHNS has been conducted eight times using the multi-stage and grouped sampling method. After 1997, the surveys were conducted in nine provinces and autonomous regions. In 2009, the rural and urban samples included 8,028 persons and 3,456 persons. During 1997–2009, the sample size of each survey was almost similar and too far lower than that in NBS.

In 2011, the Chinese Household Finance Survey (CHFS) by the Southwestern University of Finance and Economics adopted a hierarchical, three-step sampling design. The sample data included 2,585 counties and/or cities in 25 provinces and autonomous regions, 8,438 households. However, Shi and Wang [25] questioned the representativeness of the sample.

Other data sources include the China Health and Retirement Longitudinal Survey (CHARLS) and Chinese Family Panel Studies (CFPS). CHARLS is large-scale project organized by the research center at China Economy of Peking University. It was designed to provide basic data for academic research on China's aging population as well as formulate and improve China's social security policy. In 2008, a preliminary survey was conducted in both the urban and rural areas of the Gansu and Zhejiang Province. Next, CFPS was designed by those at the research center of Peking University to trace and gather three levels of data: individual, family, and community. In 2007, two test surveys with 140 households were completed in Beijing, Hebei, and Shanghai. In 2008, the exploratory studies were conducted in Beijing, Shanghai, and Guangdong and in the following year, the instrumental test panel studies were performed in the three cities.

In sum, the *China Statistics Yearbook* has several issues, while NBS has a large number of samples that have wide coverage and can be traced back to the beginning of China's reform and opening up. Thus, NBS would be an ideal data source to analyze residents' income distribution in China. Unfortunately, data provided by NBS are only from grouped households; in other words, the current grouped data include mean values that neglect income inequality within the group and consequently underestimate income disparity [11]. Moreover, the NBS survey data should be accessible to all so that great strides can be made in studies on income inequality in China.

Without access to NBS' original data, some studies calculated the Gini coefficient using China's per capita GDP [26] or per capita NI [27]. Even if per capita GDP and NI were closely related with per capita income, the Gini coefficients calculated from GDP or NI were incomparable with that obtained from income. Kanbur and Zhang [28] calculated the Gini coefficient using provincial per capita consumption and emphasized the limitations of data in measuring income inequality. Some scholars have even tried to break through the data insufficiency bottleneck by, for example, employing new statistical methods to recover missing information. Using *China Statistics Yearbook's* grouped data, Wu and Perloff [29] restored income distribution for all residents for 1985–2001 and revealed an income inequality in China. Using the same method, Wang [30] studied income distribution in China's urban and rural areas and Chi et al. [31] analyzed the income distribution in these areas from 1987 to 2004. In short, these studies employed nonparametric estimation methods. However, Zhang et al. [32] showed that the error in nonparametric estimation may be much larger than that in the parameter estimation to fit income distribution. Therefore, to study income inequality in China, it is of great significance to explore a breakthrough in methodology and estimate overall income distribution using NBS' grouped data.

In addition, it is imperative to examine the income distribution function when estimating overall income distribution using grouped data for resident income. Studies on income distribution functions have a long history. Over a century ago, Pareto [33] proposed the Pareto distribution, which earned him the same stance as Lorenz's study on income distribution [34]. Gibrat [35] stated that log-normal distribution could be a good fit for income distribution. However, successive research indicated that this distribution would underestimate the income of high-income group [36]. Distribution functions offer a host of analytical tools for studies on income inequality and promoted remarkable development. Subsequent studies made significant contributions to the field [37–44].

In China, research on income distribution function began much later. Wang [45] used the Pareto distribution to fit income data for China from 1988 to 1995. Mao et al. [46] adopted gamma distribution to fit the income of China's urban households from 2005 to 2007. Using per capita disposable income, Wang [47] studied the income distribution of rural residents and concluded that log-normal distribution had the best fitting effect. Duan and Chen [48] suggested that the national and regional per capita income of urban and rural families obeyed the mixed distribution of the Pareto distribution, normal distribution, and exponential distribution. Huang and Liu [49] adopted the nonparametric method to fit China's income distribution. Using the same method, Wang [30] explored income distribution from 1985 to 2009. Hu et al. [8] introduced different fitting methods and fit the income of rural and urban residents with the Weibull distribution, log-normal distribution, and beta distribution of the second kind (B2). The empirical results showed that B2 enjoyed the best fitting effect, in view of which the Gini coefficient of China's resident income was calculated. Hu [9] hypothesized that the generalized beta distribution of the second kind (GB2) had the best fitting effect and estimated China's Gini coefficient from 1985 to 2009. Chen et al. [50] focused on the numerical feature of the distribution function and its application to income inequality. Zhang et al. [32] compared the fitting effects of the different distribution functions and the nonparametric estimation method and showed that a three-parameter distribution function was superior to a two-parameter distribution function, and GB2 with four parameters had the best fitting effect in the income distribution of urban residents in the Anhui Province. In addition, they believed that when analyzing a distribution with complications arising from limited parameters, the parameter estimation method was clearly less capable than the nonparametric one in adjusting the distribution shape. However, an analysis of the smooth unimodal density distribution revealed that too many parameters in the nonparametric methods produced redundant information, in other words, "noise" that influenced the fitting effect. Thus, the parameter method is superior to the nonparametric one.

Various distribution functions were employed in examining the income distribution to determine the distribution function with the highest goodness of fit. This is due to the dissimilar features of income distribution in various

countries and regions during different periods; that is, no distribution function was universal. For example, Tachibanaki et al. [51] employed six commonly used distribution functions to research the income distribution of residents in Japan. McDonald [42] and McDonald and Xu [52] analyzed US household income for 1970, 1975, 1980, and 1985 and compared the fitting effects of 11 distribution functions. McDonald and Mantrala [53] adopted 15 types of distribution functions to analyze US household income for 1970, 1975, 1980, 1985, 1990, and 1995. Hu et al. [8] used grouped data and compared the fitting effects of three distribution functions. Using microeconomic survey data on urban households and three nonparametric methods, Zhang et al. [32] compared nine types of distribution functions to analyze the goodness of fit.

Drawing on the above, this study uses 20 sets of grouped income data on urban and rural residents by the *China Yearbook of Household Survey, 2013*, and eight types of distribution functions to fit the income distribution. Accordingly, we calculated the Gini coefficient in China and its changing tendency. This study makes the following theoretical and practical contributions.

(1) The *China Yearbook of Household Survey, 2013*, has published 20 sets of grouped income data, which is the most scientific and accurate compared to its previous publications. The 20 sets of grouped data provide a larger amount of information than the previous seven sets of grouped data and a wider income range of rural residents than the *China Rural Survey Yearbook*. Thus, estimating the income distribution fit using this data source is more reliable.

(2) The paper compares eight distribution functions, of which GB2, B2, Singh-Maddala (SM), and Dagum distribution are multiple parameter distribution functions. In comparison with the two-parameter distribution function, the multiple parameter distribution function has more parameters that have a greater effect on a function's shape; thus, the latter has stronger control over distribution shape and enjoys a better fitting effect. Moreover, to the authors' knowledge, no study has applied GB2, B2, and Dagum distribution to study the income distribution of urban and rural residents in China. We attach more importance to the fitting effect of GB2 because the goodness of fit is very high when GB2 is adopted to examine the income distribution of overseas residents. To this effect, Zhang et al. [32] show that GB2 has the best fitting effect among nine distribution functions and three nonparametric methods. However, the question remains whether GB2 is a good fit to the income distribution of Chinese rural and urban residents.

(3) When the distribution function fits the actual income distribution, the fitting effect of the two ends of distribution—low- and high-income group—is not good enough. To this effect, we focus on the goodness of fit of the two ends when comparing the goodness of fit of the different distribution functions.

(4) Against the background of urbanization, we fit different urbanization rates using 2012 data and estimate the contribution rate of the intrarural and intraurban Gini coefficients and the Gini between rural and urban areas to that of

the overall residents in China; this can help in framing policies aimed at narrowing the income gap in China.

The remainder of this paper is organized as following. Section 2 introduces the data and method. Section 3 compares the fitting effects of the eight distribution functions on income distribution. Section 4 explores income distribution, China's Gini coefficient, and the influence of urbanization on the Gini coefficient. Section 5 provides suggestions and concludes the paper.

2. Data and Method

2.1. Data. As mentioned, the overaggregated statistical data in the *Chinese Statistical Yearbook* and *China Yearbook of Rural Household Survey* hinder careful and in-depth analyses of income inequality. Whereas the *Chinese Statistical Yearbook* provides seven sets of grouped income data for urban residents, the *China Yearbook of Rural Household Survey* offers 20 sets of grouped income data for rural residents; however, the grouping of high-income groups is extensive. For instance, the *China Yearbook of Rural Household Survey, 2011*, presented 20 sets of grouped data in which rural households with per capita net income exceeding RMB5,000 accounted for 52.41%; but the yearbook did not subdivide such households.

Furthermore, the 20 sets of grouped income data by the *China Urban Life and Price Yearbook* have only been updated until 2010. Moreover, the data were based on per capita income and not on income per household. Since 2012, the *China Yearbook of Rural Household Survey* and *China Urban Life and Price Yearbook* merged into the *China Yearbook of Household Survey*.

In 2012, the *China Yearbook of Household Survey* issued 20 new sets of grouped income data on rural residents. The annual household income was grouped by less than RMB100 and more than RMB5,000, extending to less than RMB2,000 and more than RMB20,000. The yearbook also published 20 sets of grouped income data on urban residents and these data were based on the income per household. Thus, the new grouped data provided a more detailed income distribution of high-income group in rural areas. In addition, the grouped data for urban households by the *China Yearbook of Household Survey* provided more information than the *Chinese Statistical Yearbook*. Thus, the data can be used to more accurately fit the income distribution of rural and urban residents. However, the latest *China Yearbook of Household Survey, 2013*, only provided data for 2005, 2009, 2010, 2011, and 2012. As opposed to Hu et al. [8], the study analyzes urban and rural income distribution using income per household rather than per capita income per household.

We compared the 20 sets of grouped income data of rural residents between the *China Yearbook of Household Survey* and *Chinese Statistical Yearbook* and found that the former were given up to two decimal places, while the latter were given one. However, after fitting the same distribution function, the parameters obtained were not the same. Thus, on account of the former providing fuller data, grouped data by the *China Yearbook of Household Survey* were used.

2.2. Eight Distribution Functions. This paper adopts eight distribution functions; their probability density functions are as follows.

(1) Log-normal distribution is

$$f_{\text{LN}}(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{\ln^2(x/\mu)}{2\sigma^2}\right) \quad x \in (0, \infty). \quad (1)$$

(2) Gamma distribution is

$$f_{\text{GAMMA}}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-(x/\beta)} \quad x \in [0, \infty). \quad (2)$$

(3) Log-logistic distribution is

$$f_{\text{LG}}(x) = \frac{\beta}{\alpha} \frac{(x/\alpha)^{\beta-1}}{[1 + (x/\alpha)^\beta]^2} \quad x \in (0, \infty). \quad (3)$$

(4) Weibull distribution is

$$f_{\text{WEIBULL}}(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^\alpha} \quad x \in (x_m, \infty). \quad (4)$$

(5) Singh-Maddala distribution is

$$f_{\text{SM}}(x) = \frac{aqx^{a-1}}{b^a [1 + (x/b)^a]^{1+q}} \quad x \in (x_m, \infty). \quad (5)$$

(6) Dagum distribution is

$$f_{\text{DAGUM}}(x) = \frac{apx^{ap-1}}{b^{ap} [1 + (x/b)^a]^{p+1}} \quad x \in (x_m, \infty). \quad (6)$$

(7) Beta distribution of the second kind (B2) is

$$f_{\text{B2}} = \frac{x^{p-1}}{b^p B(p, q) [1 + (x/b)]^{p+q}} \quad x \in (0, \infty). \quad (7)$$

(8) Generalized beta distribution of the second kind (GB2) is

$$f_{\text{GB2}}(x) = \frac{ax^{ap-1}}{b^{ap} B(p, q) [1 + (x/b)^a]^{p+q}} \quad x \in (x_m, \infty). \quad (8)$$

The log-normal, gamma, log-logistic, and Weibull distributions are two-parameter distribution functions; the two parameters are scale and shape. This paper focuses on the parameters and interrelationships of multiple parameter distributions. b is a scale parameter in GB2 and a , p , and q are shape parameters. In SM, b is a scale parameter, and a and q are shape parameters. q only affects the right end and a affects both. When $p = 1$, GB2 yields the SM distribution. In the Dagum distribution, b is a scale parameter and a and p are shape parameters. When $q = 1$, GB2 becomes the Dagum distribution. In the B2 distribution, b is a scale parameter, and

p and q are shape parameters. When $a = 1$, GB2 yields the B2 distribution.

Zhang et al. [32] compared the Lomax distribution with that of Pareto. More precisely, the Lomax distribution applies to income distributions with a significantly wide income gap (the Gini coefficient is at least more than 0.5) [43]. Pareto [33] believed that all income distributions obeyed the Pareto distribution. However, empirical studies state that the Pareto distribution was not a good fit for the distribution of overall income groups but only applied to the income distribution of 1–3% of the highest income group [54]. Therefore, this study adopts B2 and not these two types of distribution. Similarly, Chotikapanich et al. [6] and Hu et al. [8] adopted B2 to analyze the income distribution of urban and rural residents and obtained a good fitting effect. However, Hu [9] posited that GB2 enjoyed the best fitting effect. Thus, this study not only compares the goodness of fit of B2 to that of the other seven distribution functions but also empirically examines whether GB2 is a good fit for the income distribution of urban and rural residents.

2.3. Parameter Estimation of Distribution Functions. The two most popular methods for estimating parameters of distribution functions are maximum likelihood estimation (MLE) and moment method estimation (MME). To apply MME, we must have complete individual data; however, according to Chotikapanich et al. [44], Hu et al. [8], since the yearbook only provides the maximum and minimum of the grouped income rather than the average, MLE would be more appropriate than MME.

To find the maximum of the likelihood function, one may use the first order condition; that is, we set the partial derivatives to zero and solve. This is feasible if the likelihood function is simple; however, the likelihood functions we encounter are oftentimes too complex to be differentiated analytically. Even if the partial derivatives can be calculated, the separate problem of solving for the zeros may still be too formidable a task. This issue has been addressed by some research for particular distributions, but the problem considered in this paper is more complex.

This study shows that the higher the number of parameters, the greater the difficulty to estimate. Here, taking GB2 as an example (see formula (8)), the log-likelihood function with n number of random samples is expressed as

$$\begin{aligned} \log L = & n \log \Gamma(p+q) + n \log a + (ap-1) \sum_{i=1}^n \log x_i \\ & - nap \log b - n \log \Gamma(p) - n \log \Gamma(q) \\ & - (p+q) \sum_{i=1}^n \log \left[1 + \left(\frac{x_i}{b} \right)^a \right], \end{aligned} \quad (9)$$

where a , b , p , and q are the partial derivatives, and thus

$$p \sum_{i=1}^n \log \left(\frac{x_i}{b} \right) + \frac{n}{a} = (p+q) \sum_{i=1}^n \log \left(\frac{x_i}{b} \right) \left[1 + \left(\frac{b}{x_i} \right)^a \right]^{-1}, \quad (10)$$

$$np = (p+q) \sum_{i=1}^n \left[\left(\frac{b}{x_i} \right)^a + 1 \right]^{-1}, \quad (11)$$

$$\begin{aligned} n\psi(p+q) + a \sum_{i=1}^n \log \left(\frac{x_i}{b} \right) \\ = n\psi(p) + \sum_{i=1}^n \log \left[1 + \left(\frac{x_i}{b} \right)^a \right], \end{aligned} \quad (12)$$

$$n\psi(p+q) = n\psi(q) + \sum_{i=1}^n \log \left[1 + \left(\frac{x_i}{b} \right)^a \right], \quad (13)$$

where $\psi(x)$ is the derivative of the gamma function's natural logarithm and $\psi(x)$ can be obtained as

$$\psi(x) = \frac{d \ln \Gamma(x)}{dx} = \frac{\Gamma'(x)}{\Gamma(x)}, \quad (14)$$

$$\text{where } \Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Formulae (13) and (15) can be calculated as

$$\begin{aligned} p &= \frac{q \sum_{i=1}^n \left[(b/x_i)^a + 1 \right]^{-1}}{n - \sum_{i=1}^n \left[(b/x_i)^a + 1 \right]^{-1}}, \\ q &= \frac{p \left\{ n - \sum_{i=1}^n \left[(b/x_i)^a + 1 \right]^{-1} \right\}}{\sum_{i=1}^n \left[(b/x_i)^a + 1 \right]^{-1}}. \end{aligned} \quad (15)$$

By substituting formulae (15) in formula (10), p and q can be expressed by including a and b . p and q are then used in formulae (12) and (13) and a and b are estimated using the Newton-Raphson iteration.

The above method is based on the assumption that we have complete individual data. The data of the present research is grouped data, so the use of this approach would not be appropriate.

To approach the technical issue of finding the parameters which maximize the likelihood function, we have several options. The most direct approach is that after defining the likelihood function, we use available software to find the parameters that maximize the value of this function. Most software such as MATLAB has a built-in command to implement this search. Our numerical experiments suggest that this method is quite efficient and robust when we have no more than 3 parameters; however, with more parameters, the likelihood function could be very irregular. To the authors' knowledge, most software uses a recursive method to search for the maximum value, which depends heavily on the initial point. On the other hand, since the 4-parameter distributions are so complicated, it is often impossible for us to choose a starting point that is "reasonably" close to the true maximum. If we fail to have a good initial point to start with, we may end up with two results: the software may fail to find a point before it runs out of the maximum number of iterations, or we obtain a local maximum instead of the global maximum.

To avoid this, we utilized the grid algorithm to find the maximum of the likelihood function. This method does not

require us to give a good starting point and has shown to be quite efficient according to our analysis. The errors between the theoretical distribution and empirical distribution, which is measured by chi-square value, are rather small. From Table 1, we see that the result is quite acceptable. It should also be mentioned that, in household income analysis, it is most important to correctly estimate the distribution of the incomes in the lowest and the highest group; for example, setting the poverty line is based on the distribution of the lowest income group while the personal taxation system design follows the distribution of the highest income group.

The basic idea of parameter estimation is explained by considering an example with three-parameter distribution functions denoted by S_1 , S_2 , and S_3 . First, using the relevant literature, the value range of S_j is estimated as $[m_j, M_j]$; thus, the extreme value point is in the cuboid $[m_1, M_1] \times [m_2, M_2] \times [m_3, M_3]$. Next, we divide the value range of each parameter into 10 parts; that is, the cuboid is divided into 1000 small cuboids. For each small cuboid, we calculate the average of the likelihood function of eight vertexes. Then, we identify the small cuboid with the maximum average value of the likelihood function. Suppose that the likelihood function is smooth enough; we have reason to believe that the extreme value point is in this small cuboid. We continue the process of further dividing the small cuboid into smaller ones and repeat the previous steps to find the smaller cuboid with the maximum average value of the likelihood function. This procedure is repeated until the dimensions of the cuboid reach our accuracy requirement. Finally, we can believe that the extreme value point is located in the center of the small cuboid.

Using the grid method, we measure the parameters of the distribution function. We judge the goodness of fit using the chi-square (χ^2) value. The χ^2 test is commonly used to test fitting degree, in which the domain is divided into several intervals. We select the χ^2 test because of the grouped data. Then, the theoretical and actual observation value in each interval is compared and checked to identify whether the error is higher than the threshold value to determine the acceptance to or rejection of the distribution function. As mentioned above, when fitting the actual income distribution with the distribution function, we may not often get a good enough fitting effect on the two ends of distribution. To test the fitting effect of different distribution functions on the two ends, we compare the actual share of population in the two ends with the theoretical one, calculated using the distribution function. In fact, the smaller the absolute value of difference, the better the fitting effect. To the best of the authors' knowledge, no similar literature has yet been found. Thus, because this research focuses on the distribution of low- and high-income groups, this method is more appropriate to study income disparity.

3. Actual Fitting Effect of the Eight Distribution Functions

Using the *China Yearbook of Household Survey, 2013*, data, we calculate the parameters and goodness of fit of the eight

distribution functions for the income distribution of rural and urban residents in China (Table 1).

Table 1 shows that the fitting effect of the three-parameter distribution function is better than that of the two-parameter distribution function, and GB2 with four parameters has the best fitting effect. In the two-parameter distribution function, the fitting effect of the log-logistic distribution is better than that of the others. As is shown, in most years, the χ^2 value and the fitting errors of the two ends are smaller than those of the other distributions. However, some scholars prefer the log-normal distribution; for instance, Souma [55] examined resident income in Japan and indicated that the log-logistic distribution is a universal structure of income distribution. By contrast, Battistin et al. [56] stated that income includes permanent income and a small amount of random income and that while the former obeys the log-normal distribution, the latter does not always do so. To this effect, the log-normal distribution fails to effectively describe income distribution.

To understand the difference between the log-normal, log-logistic, and actual income distribution, we adopt survey data on about 12,000 urban households in the Sichuan Province for 2008 and compare the differences between the theoretical and practical values of the two distribution functions mentioned above.

In Figure 1, the x -axis denotes the income quantile and the y -axis is the difference between the theoretical and practical value (estimation error). If the difference is close to 0, the fitting effect is good. On the left is the log-normal distribution and to the right is the log-logistic distribution.

Figure 1 demonstrates the difference between the theoretical and practical value fitted by the log-normal distribution gathered in the high-income group, and the negative difference shows that the log-normal distribution function underestimates the income of this group, which is similar to the findings of Zhang et al. [32]. They believed that the fast convergence rate of the ends of the log-normal distribution often led to the underestimation of the high-income group. By contrast, the fitting effect of the log-logistic distribution is better, that is, the difference between the theoretical and practical value is not obvious, which shows that there is no significant underestimation of income in the high-income group.

In the three-parameter distributions, the fitting effect of the Dagum distribution on the low-income group is better than that of the SM distribution; however, the fitting effect even on the high-income groups is often not as good as that of the SM distribution. For the urban residents, the Dagum distribution has a better fitting effect than that of the SM distribution, whereas the opposite holds true for the rural residents. Overall, these two distributions are comparable. There is no major difference in the fitting effect between the B2 distribution function and the SM and Dagum distributions when B2 fits the income distribution of the urban residents; however, B2 is more effective than the latter two for rural residents.

As for GB2, the χ^2 value and the fitting error of the two ends are smaller than that of other distributions, which is similar to the conclusions of McDonald [42] and Zhang et al. [32]. Like before, the GB2 distribution has four parameters, of

TABLE 1: Fitting results of the eight distribution functions.

(a)										
	Year	mu	Sigma	chi ²	Theoretical Pr ($X < 10000$)	Actual Pr ($X < 10000$)	Error	Theoretical Pr ($X > 100000$)	Actual Pr ($X > 100000$)	Error
Log-normal (urban)	2005	0.7803	0.5644	3893	0.0000	0.0091	0.0091	0.0709	0.0043	0.0666
	2009	0.4023	0.6054	406	0.0004	0.0018	0.0014	0.0231	0.0245	0.0014
	2010	0.5113	0.5942	658	0.0002	0.0012	0.0010	0.0323	0.0043	0.0280
	2011	0.6460	0.5861	1989	0.0001	0.0013	0.0012	0.0501	0.0511	0.0010
	2012	0.7803	0.5644	3893	0.0000	0.0008	0.0008	0.0709	0.0702	0.0007
(rural)	2005	-1.2572	0.6703	298	0.2996	0.2943	0.0053	0.0018	0.0037	0.0019
	2009	-0.8127	0.6994	124	0.1273	0.1329	0.0056	0.0157	0.0157	0.0000
	2010	-0.6697	0.6935	198	0.0877	0.0952	0.0075	0.0247	0.0235	0.0012
	2011	-0.5066	0.7157	600	0.0617	0.074	0.0123	0.0468	0.0408	0.0060
	2012	-0.3812	0.7240	614	0.0449	0.0563	0.0114	0.0689	0.0604	0.0085
(b)										
	Year	Alpha	Beta	chi ²	Theoretical Pr ($X < 10000$)	Actual Pr ($X < 10000$)	Error	Theoretical Pr ($X > 100000$)	Actual Pr ($X > 100000$)	Error
Weibull (urban)	2005	1.6159	12252	21556	0.0520	0.0091	0.0429	0.0001	0.0043	0.0042
	2009	1.7462	20015	8004	0.0178	0.0018	0.0160	0.0071	0.0245	0.0174
	2010	1.8101	22119	7053	0.0128	0.0012	0.0116	0.0126	0.0043	0.0083
	2011	1.8754	25077	6374	0.0087	0.0013	0.0074	0.0260	0.0511	0.0251
	2012	1.9963	28232	5361	0.0051	0.0008	0.0043	0.0437	0.0702	0.0265
(rural)	2005	1.3302	3803	12461	0.3464	0.2943	0.0521	0.0001	0.0037	0.0036
	2009	1.4329	6141	5288	0.1816	0.1329	0.0487	0.0044	0.0157	0.0113
	2010	1.4913	7075	4811	0.1410	0.0952	0.0458	0.0090	0.0235	0.0145
	2011	1.5123	8366	3471	0.1085	0.074	0.0345	0.0238	0.0408	0.0170
	2012	1.5399	9469	2991	0.0872	0.0563	0.0309	0.0423	0.0604	0.0181
(c)										
	Year	Alpha	Beta	chi ²	Theoretical Pr ($X < 10000$)	Actual Pr ($X < 10000$)	Error	Theoretical Pr ($X > 100000$)	Actual Pr ($X > 100000$)	Error
Gamma (urban)	2005	1.9364	0.1817	7378	0.3201	0.0091	0.3110	0.0002	0.0043	0.0041
	2009	2.0807	0.2681	2471	0.1545	0.0018	0.1527	0.0056	0.0245	0.0189
	2010	2.2005	0.2905	2010	0.1144	0.0012	0.1132	0.0111	0.0043	0.0068
	2011	2.1762	0.3473	1220	0.0864	0.0013	0.0851	0.0275	0.0511	0.0236
	2012	2.2020	0.3883	960	0.0675	0.0008	0.0667	0.0468	0.0702	0.0234
(rural)	2005	2.7728	0.3934	7135	0.0233	0.2943	0.2710	0.0002	0.0037	0.0035
	2009	2.9875	0.5948	2972	0.0051	0.1329	0.1278	0.0098	0.0157	0.0059
	2010	3.1397	0.6252	2420	0.0031	0.0952	0.0921	0.0165	0.0235	0.0070
	2011	3.2777	0.6794	2155	0.0017	0.074	0.0723	0.0312	0.0408	0.0096
	2012	3.5778	0.7013	1581	0.0007	0.0563	0.0556	0.0506	0.0604	0.0098
(d)										
	Year	Alpha	Beta	chi ²	Theoretical Pr ($X < 10000$)	Actual Pr ($X < 10000$)	Error	Theoretical Pr ($X > 100000$)	Actual Pr ($X > 100000$)	Error
Log-logistic (urban)	2005	0.9040	2.8806	204	0.0128	0.0091	0.0037	0.0072	0.0043	0.0029
	2009	1.4996	2.8939	294	0.0029	0.0018	0.0011	0.0297	0.0245	0.0052
	2010	1.6751	2.9533	179	0.0019	0.0012	0.0007	0.0381	0.0043	0.0338
	2011	1.9143	2.9914	122	0.0012	0.0013	0.0001	0.0536	0.0511	0.0025
	2012	2.1913	3.0954	165	0.0006	0.0008	0.0002	0.0722	0.0702	0.0020

(d) Continued.

	Year	Alpha	Beta	chi ²	Theoretical Pr ($X < 10000$)	Actual Pr ($X < 10000$)	Error	Theoretical Pr ($X > 100000$)	Actual Pr ($X > 100000$)	Error
(rural)	2005	0.2833	2.5823	179	0.2892	0.2943	0.0051	0.0064	0.0037	0.0027
	2009	0.4450	2.4415	444	0.1243	0.1329	0.0086	0.0249	0.0157	0.0092
	2010	0.5149	2.4686	435	0.0883	0.0952	0.0069	0.0339	0.0235	0.0104
	2011	0.6097	2.3930	634	0.0649	0.074	0.0091	0.0551	0.0408	0.0143
	2012	0.6919	2.3664	537	0.0504	0.0563	0.0059	0.0750	0.0604	0.0146

(e)

	Year	a	b	q	chi ²	Theoretical Pr ($X < 10000$)	Actual Pr ($X < 10000$)	Error	Theoretical Pr ($X > 100000$)	Actual Pr ($X > 100000$)	Error
SM (urban)	2005	2.8449	0.9210	1.0358	201	0.0133	0.0091	0.0042	0.0068	0.0043	0.0025
	2009	2.8301	1.5537	1.0696	276	0.0032	0.0018	0.0014	0.0280	0.0245	0.0035
	2010	2.8791	1.7442	1.0815	154	0.0021	0.0012	0.0009	0.0358	0.0043	0.0315
	2011	2.9901	1.9156	1.0013	122	0.0012	0.0013	0.0001	0.0535	0.0511	0.0024
	2012	3.1032	2.1829	0.9921	166	0.0006	0.0008	0.0002	0.0725	0.0702	0.0023
(rural)	2005	2.3325	0.3281	1.2774	69	0.2953	0.2943	0.0010	0.0045	0.0037	0.0008
	2009	2.1483	0.5739	1.4771	68	0.1358	0.1329	0.0029	0.0173	0.0157	0.0016
	2010	2.1951	0.6540	1.4516	71	0.0987	0.0952	0.0035	0.0252	0.0235	0.0017
	2011	2.0900	0.8325	1.5984	55	0.0761	0.074	0.0021	0.0422	0.0408	0.0014
	2012	2.0976	0.9212	1.5373	78	0.0594	0.0563	0.0031	0.0623	0.0604	0.0019

(f)

	Year	a	b	p	chi ²	Theoretical Pr ($X < 10000$)	Actual Pr ($X < 10000$)	Error	Theoretical Pr ($X > 100000$)	Actual Pr ($X > 100000$)	Error
Dagum (urban)	2005	2.8228	0.8751	1.0633	192	0.0117	0.0091	0.0026	0.0077	0.0043	0.0034
	2009	2.8615	1.4742	1.0330	291	0.0027	0.0018	0.0009	0.0304	0.0245	0.0059
	2010	2.9602	1.6806	0.9935	179	0.0019	0.0012	0.0007	0.0379	0.0043	0.0336
	2011	2.9436	1.8707	1.0465	117	0.0010	0.0013	0.0003	0.0548	0.0511	0.0037
	2012	3.0301	2.1284	1.0611	158	0.0005	0.0008	0.0003	0.0740	0.0702	0.0038
(rural)	2005	2.7744	0.3311	0.7543	99	0.2948	0.2943	0.0005	0.0051	0.0037	0.0014
	2009	2.7253	0.5387	0.7225	186	0.1356	0.1329	0.0027	0.0198	0.0157	0.0041
	2010	2.7362	0.6102	0.7476	194	0.0986	0.0952	0.0034	0.0281	0.0235	0.0046
	2011	2.7363	0.7516	0.7027	187	0.0770	0.074	0.0030	0.0456	0.0408	0.0048
	2012	2.6556	0.8286	0.7419	215	0.0598	0.0563	0.0035	0.0660	0.0604	0.0056

(g)

	Year	p	b	q	chi ²	Theoretical Pr ($X < 10000$)	Actual Pr ($X < 10000$)	Error	Theoretical Pr ($X > 100000$)	Actual Pr ($X > 100000$)	Error
B2 (urban)	2005	6.0780	0.8234	5.5730	84	0.0074	0.0091	0.0017	0.0037	0.0043	0.0006
	2009	5.4752	1.7844	6.4401	104	0.0010	0.0018	0.0008	0.0227	0.0245	0.0018
	2010	5.4025	2.2509	7.1251	172	0.0006	0.0012	0.0006	0.0304	0.0334	0.0030
	2011	5.5332	2.5981	7.3618	429	0.0003	0.0013	0.0010	0.0472	0.0511	0.0039
	2012	5.7485	3.2562	8.3442	499	0.0001	0.0008	0.0007	0.0665	0.0702	0.0037

(g) Continued.

	Year	p	b	q	chi^2	Theoretical Pr	Actual Pr	Error	Theoretical Pr	Actual Pr	Error
						($X < 10000$)	($X < 10000$)		($X > 100000$)	($X > 100000$)	
(rural)	2005	6.3929	0.1824	4.2436	55	0.2937	0.2943	0.0006	0.0034	0.0037	0.0003
	2009	3.6744	0.6920	5.4969	29	0.1317	0.1329	0.0012	0.0150	0.0157	0.0007
	2010	3.6285	0.8812	5.9331	29	0.0939	0.0952	0.0013	0.0225	0.0235	0.0010
	2011	3.0843	1.5142	7.0842	66	0.0720	0.0740	0.0020	0.0395	0.0408	0.0013
	2012	3.0338	1.7731	7.1678	28	0.0551	0.0563	0.0012	0.0597	0.0604	0.0007

(h)

	Year	a	p	b	q	χ^2	Theoretical Pr ($X < 10000$)	Actual Pr ($X < 10000$)	Error	Theoretical Pr ($X > 100000$)	Actual Pr ($X > 100000$)	Error
GB2 (urban)	2005	1.6573	2.4300	8779	2.3331	20	0.0090	0.0091	0.0001	0.0048	0.0043	0.0005
	2009	1.6167	2.4306	15763	2.6036	31	0.0015	0.0018	0.0003	0.0246	0.0245	0.0001
	2010	1.9321	1.8440	17786	2.0260	29	0.0012	0.0012	0.0000	0.0333	0.0334	0.0001
	2011	2.2615	1.5181	19318	1.5423	61	0.0007	0.0013	0.0006	0.0515	0.0511	0.0004
	2012	2.2335	1.6316	22226	1.6731	99	0.0003	0.0008	0.0005	0.0703	0.0702	0.0001
(rural)	2005	1.5553	2.2614	2805	2.2292	46	0.2942	0.2943	0.0001	0.0039	0.0037	0.0002
	2009	1.4365	1.9452	6088	2.8369	12	0.1328	0.1329	0.0001	0.0158	0.0157	0.0001
	2010	1.4720	1.8920	7136	2.8269	7	0.0953	0.0952	0.0001	0.0235	0.0235	0.0000
	2011	1.6115	1.4737	9280	2.5437	24	0.0739	0.0740	0.0001	0.0410	0.0408	0.0002
	2012	1.3855	1.8458	11586	3.3780	14	0.0562	0.0563	0.0001	0.0606	0.0604	0.0002

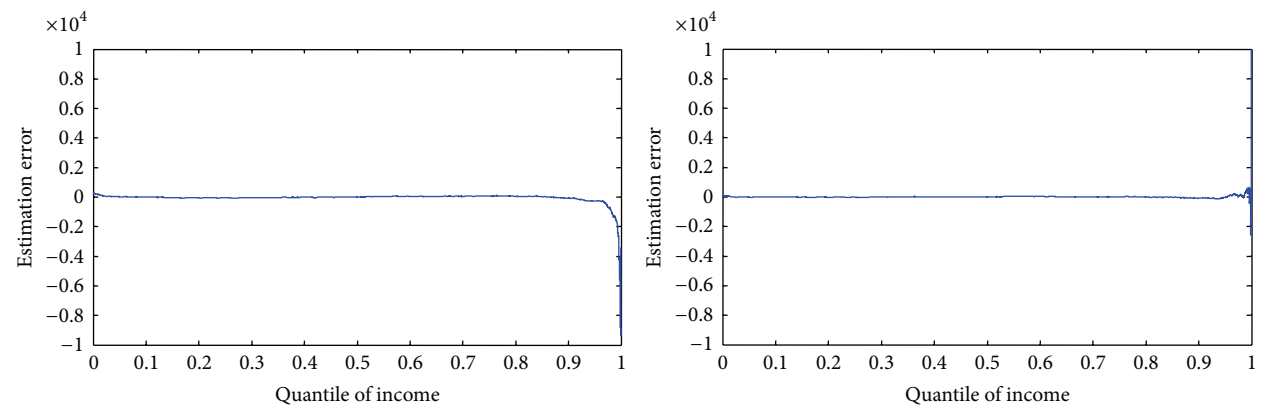


FIGURE 1: Difference between practical and theoretical values fitted by log-logistic and log-normal distributions to resident income in Sichuan in 2008.

which three are shape parameters. Therefore, GB2 has a stronger ability to adjust and control shape than the two- and three-parameter distribution functions. Thus, GB2 enjoys the best fitting effect on the income distribution of rural and urban residents in the eight distribution functions. Next, we compare the Gini coefficient calculated using different distribution functions. Here, we take the example of urban residents' income distribution in 2012.

Table 2 shows that the Gini coefficient using GB2 is 0.3151, which is regarded as a standard because it is the best fitting effect of GB2. In comparison, we find that the Gini coefficient estimated by the gamma distribution is the largest (14% higher than the standard) and that by the Weibull distribution

is the smallest (7% lower than the standard). This proves that the choice of distribution function has significant influence on the calculation of the Gini coefficient.

4. Income Distribution and Gini Coefficient of Residents in China

The Gini coefficient of national residents can be calculated using the above results. According to the parameters of income distribution and the ratio of urban to rural population by the *China Statistical Yearbook*, a random number is generated using the Monte Carlo method. This random number represents the actual income distribution of national

TABLE 2: Gini coefficient of urban residents using different distribution functions in 2012.

Distribution function	Log-normal	Gamma	Log-logistic	Weibull	B2	Dagum	SM	GB2
Gini	0.3100	0.3593	0.3224	0.2933	0.3050	0.3251	0.3229	0.3151

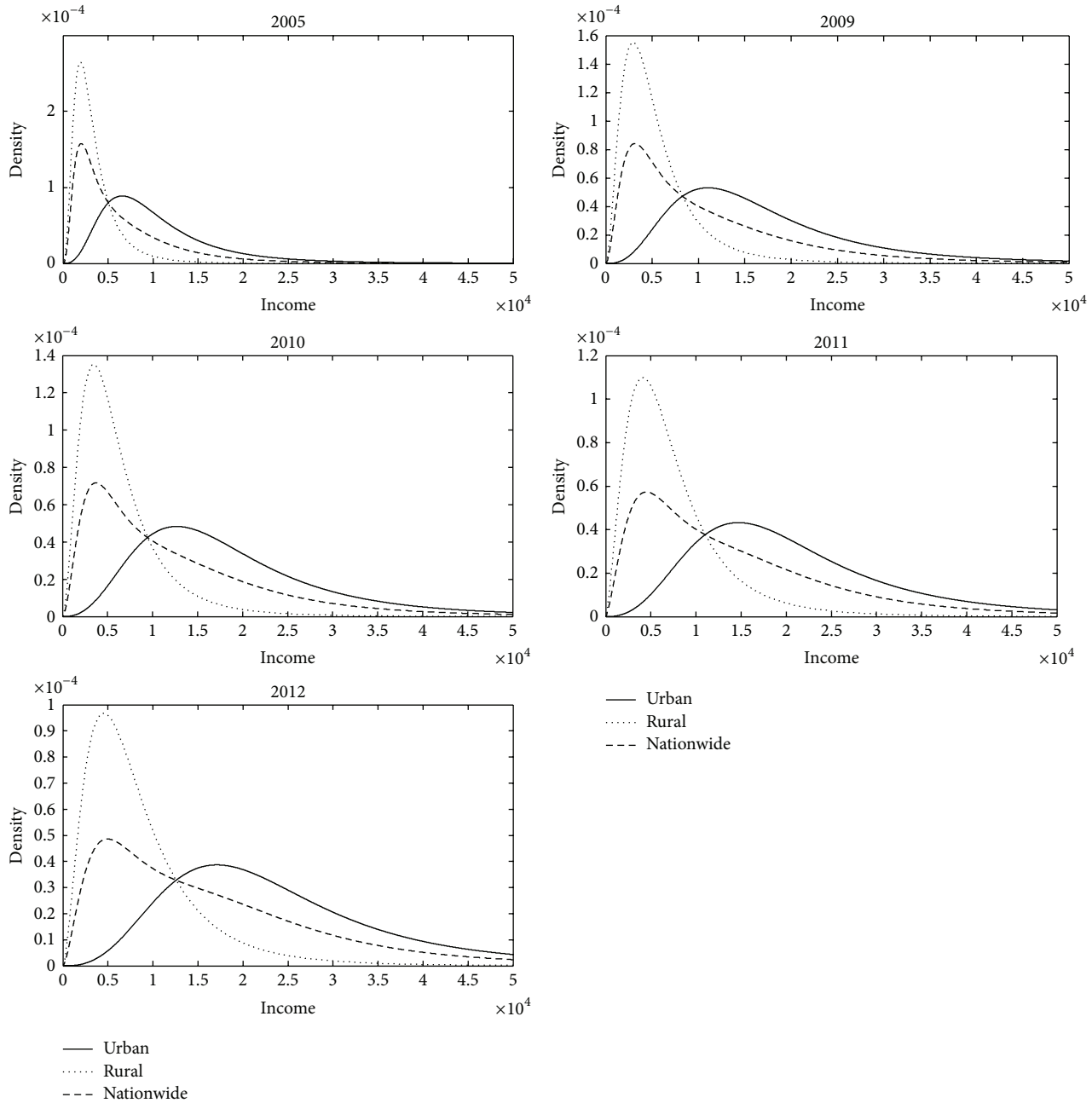


FIGURE 2: Income distribution of rural, urban, and national residents from 2005 to 2012.

residents, which helps estimate the Gini coefficient of residents in China. In this section, we explore the Gini coefficients of residents in China.

4.1. Income Distribution of Residents in China. We use the GB2 distribution function and the rural to urban population ratio from the *China Statistical Yearbook* to obtain the income

distribution of rural, urban, and national residents between 2005 and 2012 (Figure 2).

In Figure 2, the income coordinates for each year are consistent. It is evident that at the end of every year income distribution becomes increasingly dense and the peaks trend to the upper right. Due to the changes in the national income distribution shape, the fitting effects of the eight distribution

TABLE 3: χ^2 values of resident income distribution fitted by different distribution functions in 2012.

Distribution function	Log-normal	Gamma	Log-logistic	Weibull	B2	Dagum	SM	GB2
χ^2	64115	64040	3700	2117	636	1984	986	503

TABLE 4: Gini coefficient of resident income in China.

Year	The ratio of the urban population	Urban Gini coefficient	Rural Gini coefficient	1000 times of simulations	10000 times of simulations	100000 times of simulations	NBS
2005	0.4300	0.3386	0.3718	0.4512	0.4552	0.4561	0.485
2009	0.4659	0.3330	0.3812	0.4500	0.4564	0.4571	0.490
2010	0.4995	0.3272	0.3757	0.4379	0.4433	0.4440	0.481
2011	0.5127	0.3274	0.3806	0.4307	0.4373	0.4380	0.477
2012	0.5257	0.3150	0.3808	0.4225	0.4277	0.4290	0.474

Note: the ratio of the urban population is taken from the *China Statistical Yearbook—2013*.

functions become significantly weak (Table 3). For the χ^2 value, GB2 has the best fitting effect among the eight functions. In the two-parameter distribution functions, the log-logistic and Weibull distribution have a better fitting effect than the others, while among the three-parameter distribution functions, B2 enjoys the best fitting effect. We also find that the fitting effect of the three-parameter distribution function is superior to that of the two-parameter function. Next, we calculate the Gini coefficient of residents in China from 2005 to 2012.

4.2. Gini Coefficient of Chinese Residents. In line with studies on GB2 [42], the Gini coefficient can be directly derived and expressed as

$$G = \frac{2B(2p+1/a, 2q-1/a)}{pB(p, q)B(p+1/a, q-1/a)} \cdot \left\{ \frac{1}{p} {}_3F_2 \left[1, p+q, 2p+\frac{1}{a}; p+1, 2(p+q); 1 \right] - \frac{1}{p+1/a} {}_3F_2 \left[1, p+q, 2p+\frac{1}{a}; p+\frac{1}{a}+1, 2(p+q); 1 \right] \right\}, \quad (16)$$

where $B(p, q)$ is the beta distribution, in which $B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1}dt$ and $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$, and ${}_3F_2$ is the generalized hypergeometric function, expressed as

$${}_3F_2(a_1, a_2, a_3; b_1, b_2; x) = \sum_{n=0}^{\infty} \left[\frac{(a_1)_n (a_2)_n (a_3)_n}{(b_1)_n (b_2)_n} \right] \frac{x^n}{n!}, \quad |x| < 1. \quad (17)$$

Here, $(a_1)_n = \Gamma(a_1+n)/\Gamma(a_1)$.

Since calculating the Gini coefficient can be complicated, we use the distribution function with the known parameters to generate a random number using the Monte Carlo method. The random number depends on the ratio of urban to national population (Table 4), after which the Gini coefficient

is calculated using $G = (1/2n^2u) \sum_i \sum_j |y_i - y_j|$, where n is the total number of households, μ is the average income, and y_i and y_j , respectively, represent the i th and j th family income. Each random number obeying a specific distribution can be viewed as the family income data. Further research reveals that, as the number of simulations increase, the result gradually converges into a specific value. Finally, after 1,00,000 simulations, we obtain the income Gini coefficients of urban, rural, and national residents for 2005–2012 (Table 4).

Table 4 indicates that the Gini coefficient of urban and national residents has been declining from 2005 to 2012, while that of rural residents is on the rise. Compared to the Gini coefficient by NBS, our results are marginally lower but show the same trend.

The Gini coefficient for 2003–2012 by NBS was calculated using new standards, new caliber, and old data for the following reasons: (1) to determine the indicators of disposable income after the integrated household survey system was conducted, (2) to resolve the problem of rich households' income being estimated lower than actual income, and (3) to adjust the classification of rural migrant worker. As for (1), the indicator was set in accordance with the *National Economic Accounting, 2008*, by the UN Statistical Commission and the *Canberra Group Handbook of Household Income Statistics (2011)* by the UN Economic Commission for Europe. More precisely, the index of the per capita net income of rural residents is adjusted to the per capita disposable income; that is, social security expenditure is deducted from and interest is added to the income of rural migrant workers. Moreover, to standardize the caliber of per capita disposable income means to further deduct all transfer spending—excluding social security expenditure, income tax, and property expenditure, mainly mortgage interest—and increase the net rental and in-kind income. Next, to solve the issue discussed in (2), NBS studied the calibration methods adopted in domestic research. By a comparison with personal income tax, NBS calculated the weights of high-income samples, adjustment coefficient, and per capita income for the smooth calibration of the high-income group. Finally, to adjust the classification mentioned in (3), migrant workers employed for more than half a year are classified under urban population from

the current rural population to maintain consistency with the demographic classification. Next, the weights of the grouped data from the urban and rural household survey are merged in the light of the urban to national population ratio (Source: “the director of household survey office in the bureau of statistics provided the Gini coefficient of resident income,” China News Service, 19:42, Feb. 1, 2013 (<http://news.sohu.com/20130201/n365335330.shtml>)).

From the above, we believe that certain factors cause a marginally higher Gini coefficient when using NBS data because NBS adjusted the income distribution of high-income group and used various internationally standard calibration methods for trial. In addition, it selected the calibration method with detailed data sources and the maximum value of calibration and finally relied on individual income tax data. Thus, the Gini coefficient was marginally higher than our estimation. According to the existing survey, the average income of migrant workers is much higher than that of rural residents. From a statistical view point, it widens the rural-urban income gap in classifying migrant workers who migrate to urban areas to find employment. Undoubtedly, the income gap between rural and urban areas plays an important role in the income disparity of residents in China. Therefore, adjusting the statistical standard leads to a marginally higher Gini coefficient.

NBS emphasizes that the adjusted result is internationally comparable; however, we notice that the household survey in China adopts this widely varying method from other countries. The Chinese respondents are required to record their spending throughout the year, while in many other countries, a survey is administered to the respondents on income and expenditure on a weekly, biweekly, or monthly basis. The data are then multiplied by 52, 26, or 12 to obtain annual data [57]. By contrast, the average data from annual surveys in China help reduce the volatility of income and expenditure. For example, a sudden income decrease in one month balances an unexpected increase in another. Compared with other countries, China's Gini coefficient can be underestimated because of the difference in statistical method [58]. We use monthly survey data of urban residents in Anhui Province to measure the monthly and annual Gini coefficient for 2008, which shows that the annual value (112%) is lower than the monthly one (116%). Thus, from the perspective of international comparison, China's Gini coefficient has been underestimated because of the use of different statistical methods. If the annual statistical data by NBS are accurate, the current Gini coefficient is underestimated. Correspondingly, depending on the household survey data in Anhui Province, the Gini coefficient of Chinese residents may exceed 0.52. Owing to the detailed data source and calibration method with a maximum calibration quantity, the Gini coefficient of Chinese residents is less than 0.6, even when the underestimation of annual statistical data is considered.

Hu [9] adopted a similar method to calculate the Gini coefficient of Chinese residents for 1985–2009. We compared the estimators of 2005 and 2009 and found that in the same years the parameters of the GB2 distribution function are not similar to ours when we fit the income of rural and urban residents. The significant differences are in the estimation of

income distribution parameters for rural residents. Although the data structures are not identical, they are all derived from the urban and rural household survey by NBS. We adopt his results and derive identical results. In view of this, his parameters are only local optimum and not global optimal, which leads to a large difference in the intrarural Gini coefficient. While Hu [9] calculated the intrarural Gini coefficient as 0.4183 for 2005 and 0.6738 for 2009, our estimations are 0.3718 and 0.3812. We question Zhijun's result because the same Gini coefficient estimations by the Department of Employment and Income Distribution at the National Development and Reform Commission are 0.38 and 0.39 (see “Yearly Report on Resident Income in China” (page 254, 2010), edited by the Department of Employment and Income Distribution at the National Development and Reform Commission, Economic Science Press). Owing to the overestimated Gini coefficient for rural residents, that of national residents in 2009 by Hu [9] is higher than 0.54. Hu et al. [8] also used NBS data and distribution functions for 1985–2008 to estimate the Gini coefficient for residents in cities and towns as well as the country and nation and derived an intrarural Gini of only 0.3370. In addition, we compare the intrarural, intraurban, and national Gini by Hu et al. [8] and Chotikapanich et al. [6] and identify a major difference among them. While the Gini coefficient of national residents by Hu [9] is marginally high, that by Chotikapanich et al. [6] is low; Chotikapanich et al. estimated a value of only 0.2827 for 1985. Thus, choosing different distribution functions can produce prominent calculation errors. In particular, even the same distribution function can produce a significant deviation if the results are only local optimal.

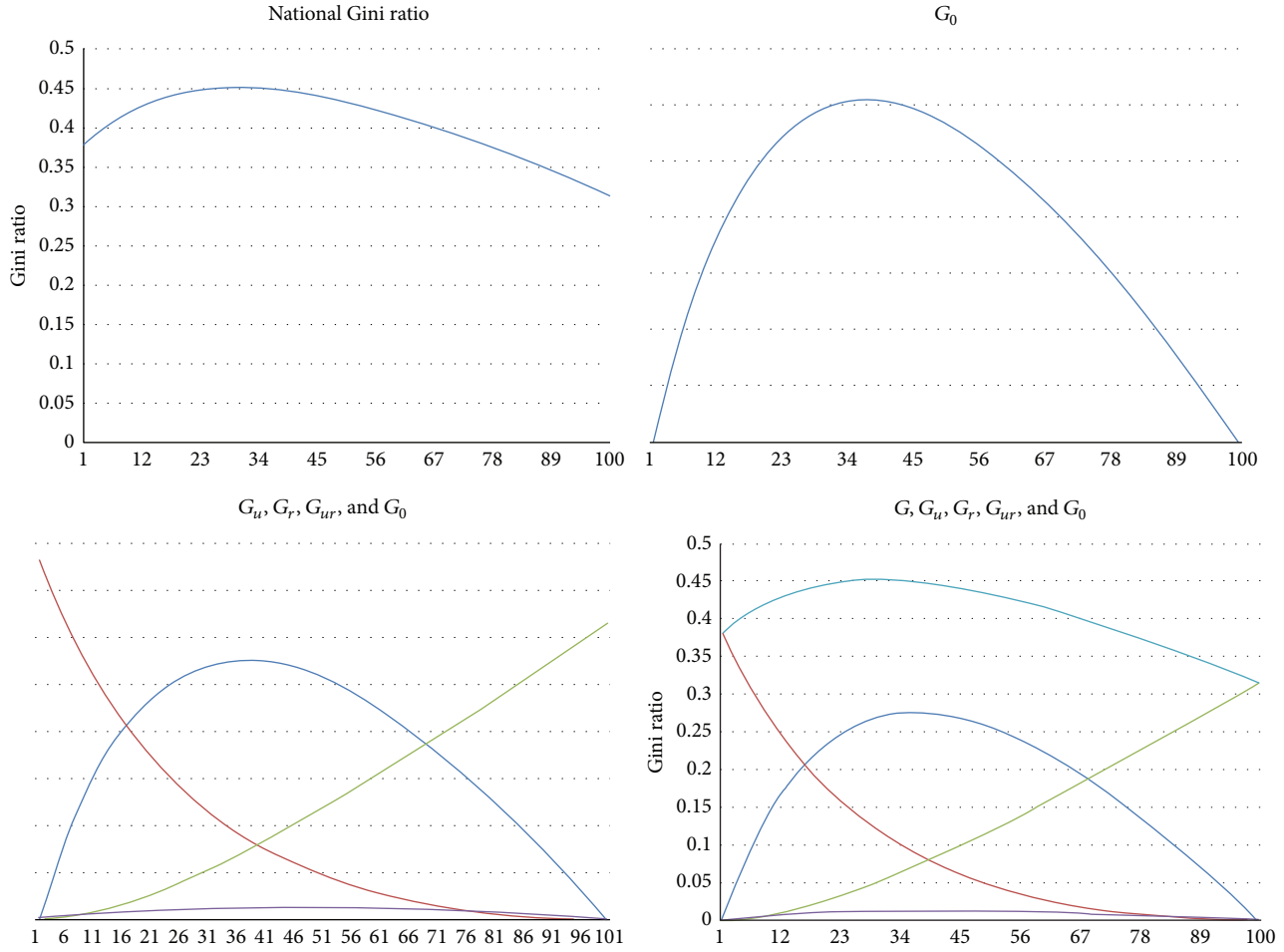
4.3. Influence of Urbanization on the Income Gap of Residents in China. According to Chen [12], urbanization can diminish the income gap in China. We examine 2012 data to verify the influence of urbanization on China's income gap. Supposing that the income distribution of rural and urban residents is fixed in 2012 but there are variations in urban to national population ratio ranging from 0% to 100%, the Gini coefficient will change from 0.3808 to 0.3150.

Using the decomposition method [59–61], the national Gini coefficient (G) can be decomposed into four parts: intrarural Gini (G_r), intraurban Gini (G_u), the Gini coefficient between the urban and rural area (G_{ur}), and the cross term of resident income in urban and rural areas (G_0):

$$G = G_{ur} + \delta G_u + \beta G_r + G_0, \quad (18)$$

where δ and β denote the results of the share of urban and rural population (P_u, P_r) multiplied by the income proportion of the urban and rural areas (I_u, I_r), which is $\delta = I_u P_u$ ($\beta = I_r P_r$). The decomposed national Gini includes four segments: G_{ur} , G_u , G_r , and G_0 . Their relevant coefficients are 1, δ , β , and 1, respectively. This formula clearly shows the contributions of each part to the national Gini coefficient.

G_0 emanates from the income overlap of urban and rural residents. Although previous studies on the influence of G_0 are not apparent, G_0 was often referred to as the residual or cross term of income, also termed the “income concentration

FIGURE 3: Changes in the national Gini: G_u , G_r , G_{ur} , and G_0 .

zone” by Bhattacharya and Mahalanobis [62]. In addition, Mookherjee and Shorrocks [63] stated that it was a frustrating interaction term that was impossible to accurately calculate. Lambert and Aronson [61] argued that the residual was a result produced by calculations for both within and between groups, presenting the overlapping degree of the income distribution of different groups, which was similar to Cowell’s [64] ideology. Lambert and Decoster’s [65] mathematical expression for G_0 was

$$G_0 = 2P_u P_r \frac{\int [1 - F_u(x)] F_r(x) dx}{\mu}. \quad (19)$$

Here, $F_u(x)$ and $F_r(x)$ are the income distribution functions of urban and rural residents and μ is the average income of residents across the country. Using formulae (18) and (19), we calculate the national Gini coefficient or the changes in G_u , G_r , G_{ur} , and G_0 because of urbanization (Figure 3; the x -axis shows the urbanization ratio).

In Figure 3, the upper right illustration shows the changes in G_0 with urbanization. Due to an extremely small effect, it is difficult to identify how it changes from the bottom left and right one. The bottom left one shows the contributions of several factors to the national income disparity in rural areas

(the $I_r P_r G_r$ curve steadily declines to the bottom right), the income inequality in urban areas (the $I_u P_u G_u$ curve steadily rises to the upper right), the income difference between urban and rural areas (G_{ur} is the inverted U curve), and G_0 (the inverted curve close to the X axis). The bottom right one is a combination of the upper left and bottom left illustrations.

From Figure 3, we conclude that urbanization can help narrow the income disparity of residents in China. With urbanization, the contribution of G_0 and that of income disparities in rural areas and between rural and urban areas to national income inequity will be in continuous decline, while the contribution of income disparity in urban areas will rise. To conduct an in-depth analysis of the influence of factors mentioned above on the national Gini coefficient, we calculate the contribution rate of various factors to the national Gini coefficient (Figure 4).

Figure 4 demonstrates that at the beginning of urbanization the contribution rate of the intrarural Gini coefficient is higher; thereafter, the income disparity between rural and urban areas makes a larger contribution. Once urbanization has been achieved the intraurban Gini coefficient plays a more important role than others. When the income distribution of urban and rural residents in 2012 is fixed,

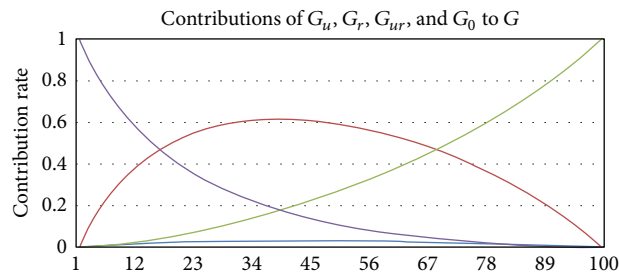


FIGURE 4: Contribution rate of G_u , G_r , G_{ur} , and G_0 to the Gini coefficient of residents in China.

the urbanization rate reaches 15.3%, and the intrarural Gini coefficient and that for the urban and rural areas makes the same contributions to the overall Gini coefficient. As for the urbanization rate in 2012, the Gini coefficient between urban and rural areas accounts for 58% of the contribution rate to the overall Gini. If the average income ratio of urban and rural residents remains unchanged when the urbanization rate is 38%, the contribution rate of the Gini coefficient for rural and urban areas to the national one reaches 61.6%. If the urbanization rate is 68.3%, the intraurban Gini coefficient and that for urban and rural areas make the same contribution to the national Gini coefficient. Currently, the income inequality between the rural and urban areas is a key factor influencing the income disparity of residents in China. Thus, accelerating urbanization can play a major role in diminishing the income inequality of residents in China.

5. Conclusions and Suggestions

Using the 20 sets of grouped rural and urban family data for 2005–2012 by the *China Yearbook of Household Survey*, this paper adopted eight common distribution functions to fit the income distribution of urban and rural residents. The results showed that GB2 has the best fitting effect among the eight distribution functions. In addition, the fitting effect of the three-parameter distribution function is superior to that of the two-parameter distribution function. In the two-parameter distribution function, the log-logistic distribution function enjoys the best fitting effect. We also compare the fitting effects of the log-logistic and log-normal distribution and reveal the reasons for the differences in fitting effect.

By referencing previous research, we describe the income distribution of residents in China. We calculate the Gini coefficient for rural and urban income as well as the overall Gini coefficient. The results show that the national Gini coefficient has been on the decline from 2005 to 2012, which also hold true for urban residents. By contrast, the overall Gini coefficient has been rising. In addition, we analyze the Gini coefficient for national residents by NBS and estimate the range of the national Gini coefficient.

To discuss the effects of urbanization on the income gap of residents in China, the parameters of income distribution function for 2012 are kept constant and the urbanization rate is assumed to range from 0% to 100%. We explore the degree of influence that the intrarural Gini, intraurban Gini, and

Gini coefficient between rural and urban areas have on the overall Gini coefficient. We conclude that the Gini coefficient between rural and urban areas makes the biggest contribution to the national Gini, which, however, appears to be on a downward trend because of urbanization. We predict that the intraurban Gini coefficient will make increasingly larger contributions and play a decisive role in the future.

In line with the above results, we make the following three suggestions. First, urbanization plays a key role in narrowing the income inequality in China and should thus be promoted. Second, higher importance should be attached to intraurban income disparity. With urbanization, the intraurban income inequality will make a larger contribution to the overall income disparity in China. Finally, fuller and more accurate data on resident income are expected to be released by NBS to strongly promote research on China's income inequality.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This paper is funded by the National Natural Science Foundation of China (Grant nos. 71473203 and 11426181), the Innovation Team Project for Basic Research Funding by the Central University (JBK120502), and the Important Basic Theory Research of Southwestern University of Finance and Economics (JBK141102).

References

- [1] Z. Chen, *Reform, Development and Income Distribution*, Fudan University Press, Beijing, China, 1999.
- [2] Z. Chen and Y. Zhou, *Rethinking Income Distribution in Reform and Development: Do Chinese Income Have a Bipolar?* Economic Science Press, Beijing, China, 2002, (Chinese).
- [3] Y. Chen, "Calculation and decomposition of the overall gini coefficient in dual economies," *Economic Research Journal*, no. 1, pp. 109–201, 2006 (Chinese).
- [4] M. Ravallion and S. Chen, "China's (uneven) progress against poverty," *Journal of Development Economics*, vol. 82, no. 1, pp. 1–42, 2007.
- [5] S. Xiang, "Calculation and regression analysis on the gini coefficient of income distribution of residents in China," *The Theory and Practice of Finance and Economics*, no. 1, pp. 75–80, 1997 (Chinese).
- [6] D. Chotikapanich, D. S. P. Rao, and K. K. Tang, "Estimating income inequality in China using grouped data and the generalized beta distribution," *Review of Income and Wealth*, vol. 53, no. 1, pp. 127–147, 2007.
- [7] Y. Xu and X. Zhang, "An estimation on a red line of the Gini coefficient in China," *Statistical Research*, no. 1, pp. 80–83, 2011 (Chinese).
- [8] Z. Hu, Z. Liu, and Z. Gong, "The estimation of the gini coefficient of overall income in China: 1985–2008," *China Economics Quarterly*, no. 4, pp. 1423–1436, 2011 (Chinese).
- [9] Z. Hu, "The estimation of the gini coefficient and social welfare based on the grouped data: 1985–2009," *The Journal*

- of Quantitative & Technical Economics*, no. 9, pp. 111–121, 2012 (Chinese).
- [10] R. Zhao, S. Li, and C. Li, *Re-examination on Income Distribution of Residents in China*, Chinese Financial & Economic Publishing House, Beijing, China, 1999, (Chinese).
 - [11] A. Khan and C. Riskin, *Inequality and Poverty in China in the Age of Globalization*, Oxford University Press, Oxford, UK, 2001.
 - [12] J. Chen, “A study on the gini coefficient of residents in China with decomposition of rural and urban areas,” *Economic Issues in China*, no. 4, pp. 33–41, 2010.
 - [13] S. Li and C. Luo, “Re-estimating the income gap between urban and rural households in China,” *The Journal of Peking University (Philosophy and Social Sciences)*, vol. 2, pp. 111–120, 2007.
 - [14] C. Fang, X. Zhang, and S. Fan, “Emergence of urban poverty and inequality in China: evidence from household survey,” *China Economic Review*, vol. 13, no. 4, pp. 430–443, 2002.
 - [15] D. Sutherland and S. Yao, “Income inequality in China over 30 years of reforms,” *Cambridge Journal of Regions, Economy and Society*, vol. 4, no. 1, pp. 91–105, 2011.
 - [16] T. Sicular, Y. Ximing, B. Gustafsson, and L. Shi, “The urban-rural income gap and inequality in China,” *Review of Income and Wealth*, vol. 53, no. 1, pp. 93–126, 2007.
 - [17] S. Li and C. Luo, “Re-estimating the income gap between urban and rural households in China,” *Procedia-Social and Behavioral Sciences*, vol. 2, no. 5, pp. 7151–7163, 2010.
 - [18] H. Wang, “Change in China’s household income and its impact on equality in the long term,” *Economic Science*, no. 2, pp. 56–67, 2005.
 - [19] S. Li, T. Shi, and B. Gustafsson, *Research on Income Distribution in China*, Beijing Normal University Press, Beijing, China, 2008 (Chinese).
 - [20] X. Shi, T. Sicular, and Y. Zhao, “Analyzing urban-rural income inequality in China,” in *Proceedings of the International Symposium on Equity and Social Justice in Transitional China*, pp. 11–12, Beijing, China, 2002.
 - [21] Z. Wei, “The effects of health on non-agricultural employment and wage determination,” *Economic Research Journal*, no. 2, pp. 64–128, 2004.
 - [22] H. Wang, “Inter-generational mobility of income distribution of residents in China,” *Economic Research Journal*, no. 1, pp. 18–25, 2005 (Chinese).
 - [23] S. Zhang, B. Xu, and Z. Hu, “The dynamic evolution and empirical test of income distribution in rural and urban areas in China,” *Statistical Research*, no. 12, pp. 32–40, 2009 (Chinese).
 - [24] N. Zhu and X. Luo, “Income growth, inequality and poverty,” *China Journal of Population Science*, no. 2, pp. 12–24, 2008.
 - [25] L. Shi and H. Wang, “To improve the credibility of estimation on gini coefficient in China—author’s discuss of China household finance survey,” *Economic Perspectives*, vol. 2, pp. 43–49, 2013.
 - [26] Z. Zhang, A. Liu, and S. Yao, “Convergence of China’s regional incomes: 1952–1997,” *China Economic Review*, vol. 12, no. 2–3, pp. 243–258, 2001.
 - [27] Y. Lin, *Empirical Research on Income Gap of Residents in China*, Peking University Press, Beijing, China, 2000 (Chinese).
 - [28] R. Kanbur and X. Zhang, “Fifty years of regional inequality in china: a journey through revolution, reform and openness,” in *Proceedings of the UNU/WIDER Project Conference on Spatial Inequality in Asia*, United Nations University Centre, Tokyo, Japan, 2003.
 - [29] X. Wu and J. M. Perloff, “China’s Income Distribution, 1995–2001,” University of California, 2005, <http://escholarship.org/uc/item/0zd6m0sf>, 2013/05/04.
 - [30] Y. Wang, “Estimation on income distribution of rural and urban residents in China: 1985–2009,” *The Journal of Quantitative & Technical Economics*, no. 6, pp. 61–73, 2012 (Chinese).
 - [31] W. Chi, B. Li, and Q. Yu, “Decomposition of the increase in earnings inequality in urban China: a distributional approach,” *China Economic Review*, vol. 22, no. 3, pp. 299–312, 2011.
 - [32] M. Zhang, J. Chen, and M. Pu, “A study on income distribution function of urban residents,” *The Journal of Quantitative & Technical Economics*, no. 3, pp. 57–71, 2013.
 - [33] V. Pareto, “La legge della domanda,” *Giornale degli Economisti*, no. 12, pp. 59–68, 1895.
 - [34] B. C. Arnold, *Pareto Distributions*, International Co-operative Publishing House, Fairland, Md, USA, 1983.
 - [35] R. Gibrat, *Les Inégalités Économiques*, Librairie du Recueil Sirey, Paris, France, 1931.
 - [36] M. Kalecki, “On the Gibrat distribution,” *Econometrica*, vol. 13, no. 2, pp. 161–170, 1945.
 - [37] R. Rutherford, “Income distributions: a new model,” *Econometrica*, vol. 23, no. 3, pp. 277–294, 1955.
 - [38] D. G. Champernowne, “A model of income distribution,” *The Economic Journal*, vol. 63, no. 250, pp. 318–351, 1953.
 - [39] J. Aitchison and J. Brown, *The Log-Normal Distribution*, Cambridge University Press, Cambridge, UK, 1957.
 - [40] S. K. Singh and G. S. Maddala, “A function for size distribution of incomes,” *Econometrica*, vol. 44, no. 5, pp. 963–970, 1976.
 - [41] C. Dagum, “A new model of personal distribution: specification and estimation,” *Economie Appliquée*, no. 30, pp. 413–437, 1977.
 - [42] J. B. McDonald, “Some generalized functions for the size distribution of income,” *Econometrica*, vol. 52, no. 3, pp. 647–663, 1984.
 - [43] C. Kleiber and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley & Sons, Hoboken, NJ, USA, 2003.
 - [44] D. Chotikapanich, W. E. Griffiths, and D. S. P. Rao, “Estimating and combining national income distributions using limited data,” *Journal of Business & Economic Statistics*, vol. 25, no. 1, pp. 97–109, 2007.
 - [45] H. Wang, “Pattern of the income distribution of residents in China—Pareto distribution,” *South China Journal of Economics*, no. 5, pp. 73–82, 2006 (Chinese).
 - [46] L. Mao, C. Hao, X. Ping, S. Liang, and H. Zhang, “International comparison of the individual income tax threshold and effect estimation on raise of minimum threshold,” *International Economic Review*, vol. 6, pp. 34–38, 2009.
 - [47] Z. Wang, “The determination of poverty level and poverty in rural areas—relying on the fitting method of income distribution function,” *Northwest Population Journal*, no. 1, pp. 1–6, 2009 (Chinese).
 - [48] J. Duan and J. Chen, “Calculation and decomposition of the regional gini coefficient based on family income distribution,” *The Journal of World Economy*, no. 1, pp. 100–122, 2010 (Chinese).
 - [49] H. Huang and L. Liu, “A method to fit income distribution function sequences and its extended application,” *Statistics & Information Forum*, no. 12, pp. 14–18, 2011 (Chinese).
 - [50] J. Chen, T. Luo, and A. Zhao, “The application of income distribution functions in the research of income equality,” *Statistical Research*, no. 9, pp. 79–86, 2013 (Chinese).

- [51] T. Tachibanaki, T. Suruga, and N. Atoda, "Estimations of income distribution parameters for individual observations by maximum likelihood method," *Journal of the Japan Statistical Society*, vol. 27, no. 2, pp. 191–203, 1997.
- [52] J. B. McDonald and Y. J. Xu, "A generalization of the beta distribution with applications," *Journal of Econometrics*, vol. 66, no. 1-2, pp. 133–152, 1995.
- [53] J. B. McDonald and A. Mantrala, "The distribution of personal income: revisited," *Journal of Applied Econometrics*, vol. 10, no. 2, pp. 201–204, 1995.
- [54] A. Harrison, "Earnings by size: a tale of two distributions," *The Review of Economic Studies*, vol. 48, no. 4, pp. 621–631, 1981.
- [55] W. Souma, "Universal structure of the personal income distribution," *Fractals*, vol. 9, no. 4, pp. 463–470, 2001.
- [56] E. Battistin, R. Blundell, and A. Lewbel, "Why is consumption more log normal than income? Gibrat's law revisited," *Journal of Political Economy*, vol. 117, no. 6, pp. 1140–1154, 2009.
- [57] J. Gibson, J. Huang, and S. Rozelle, "Why is income inequality so low in China compared to other countries?: the effect of household survey methods," *Economics Letters*, vol. 71, no. 3, pp. 329–333, 2001.
- [58] A. Deaton, "Data and econometric tools for development analysis," *Handbook of Development Economics*, vol. 3, pp. 1785–1882, 1995.
- [59] F. A. Cowell, "On the structure of additive inequality measures," *The Review of Economic Studies*, vol. 47, no. 3, pp. 521–531, 1980.
- [60] A. F. Shorrocks, "Inequality decomposition by population subgroups," *Econometrica*, vol. 52, no. 6, pp. 1369–1385, 1984.
- [61] P. J. Lambert and J. R. Aronson, "Inequality decomposition analysis and the gini coefficient revisited," *The Economic Journal*, vol. 103, no. 420, pp. 1221–1227, 1993.
- [62] N. Bhattacharya and B. Mahalanobis, "Regional disparities in household consumption in India," *Journal of the American Statistical Association*, vol. 62, no. 317, pp. 143–161, 1967.
- [63] D. Mookherjee and A. Shorrocks, "A decomposition analysis of the trend in UK income inequality," *The Economic Journal*, vol. 92, no. 368, pp. 886–902, 1982.
- [64] F. Cowell, "Measuring inequality," in *Handbook: Income Distribution*, A. Atkinson and F. Bourguignon, Eds., Elsevier-North Holland, 2000.
- [65] P. Lambert and A. Decoster, "The gini coefficient reveals more," *International Journal of Statistics*, vol. 63, pp. 373–400, 2005.

